

**IMPLEMENTASI ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES  
TERHADAP KELUHAN SARANA DAN PRASARANA DI BIRO  
ADMINISTRASI UMUM UNIVERSITAS MUHAMMADIYAH SURAKARTA**



**Disusun sebagai salah satu syarat menyelesaikan Program Studi Strata I pada  
Program Studi informatika Fakultas Komunikasi dan Informatika**

**Oleh:**

**DIMAS YOGA PRATAMA**

**L200160065**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**2020**

**HALAMAN PERSETUJUAN**

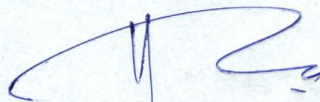
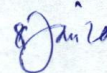
**IMPLEMENTASI ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES  
TERHADAP KELUHAN SARANA DAN PRASARANA DI BIRO  
ADMINISTRASI UMUM UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**PUBLIKASI ILMIAH**

oleh:

**DIMAS YOGA PRATAMA**  
**L200160065**

Telah diperiksa dan disetujui untuk diuji oleh:  
Dosen Pembimbing

  
**Dr., Ir. Bana Handaga, M.T.**  
NIK.793 



**HALAMAN PENGESAHAN**

**IMPLEMENTASI ANALISIS SENTIMEN MENGGUNAKAN NAIVE  
BAYES TERHADAP KELUHAN SARANA DAN PRASARANA DI BIRO  
ADMINISTRASI UMUM UNIVERSITAS MUHAMMADIYAH  
SURAKARTA**

**OLEH**  
**DIMAS YOGA PRATAMA**  
**L200160065**

Telah dipertahankan di depan Dewan Penguji  
Fakultas Komunikasi dan Informatika  
Universitas Muhammadiyah Surakarta  
Pada hari Kamis, 16 Januari 2020  
dan dinyatakan telah memenuhi syarat

**Dewan Penguji:**

1. Dr. Ir. Bana Handaga, M.T.  
(Ketua Dewan Penguji)
2. Dr. Heru Supriyono, M.Sc  
(Anggota I Dewan Penguji)
3. Aris Rakhmadi, S.T., M.Eng  
(Anggota II Dewan Penguji)

(.....)  
(.....)  
(.....)



Dekan  
Fakultas Komunikasi dan Informatika

Nurgiyatna, S.T., M.Sc., Ph.D.  
NIK 881



Ketua  
Program Studi Informatika

Heru Supriyono, S.T., M.Sc., PhD  
NIK 970

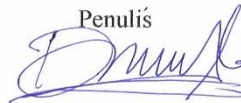
## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam naskah publikasi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

Surakarta, 29 Januari 2020

Penulis



**DIMAS YOGA PRATAMA**

**L200160065**



**UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

**SURAT KETERANGAN LULUS PLAGIASI**

**No Surat .IQ/4.4: II.3/Inf-FKI/1/2020**

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : Dimas Yoga Pratama  
NIM : **L200160065**  
Judul : **IMPLEMENTASI ANALISIS SENTIMEN MENGGUNAKAN NAIVE BAYES  
TERHADAP KELUHAN SARANA DAN PRASARANA DI BIRO  
ADMINISTRASI UMUM UNIVERSITAS MUHAMMADIYAH SURAKARTA**  
Program Studi : Informatika  
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 23 januari 2020

Biro Skripsi Informatika

**Ihsan Cahyo Utomo, S.Kom., M.Kom.**



UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

## IMPLEMENTASI ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES TERHADAP KELUHAN SARANA DAN PRASARANA DI BIRO ADMINISTRASI UMUM UNIVERSITAS MUHAMMADIYAH SURAKARTA

### ORIGINALITY REPORT

<b>19%</b>	<b>15%</b>	<b>7%</b>	<b>17%</b>
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

### PRIMARY SOURCES

<b>1</b>	<b>docplayer.info</b> Internet Source	<b>4%</b>
<b>2</b>	<b>Submitted to Universitas Brawijaya</b> Student Paper	<b>2%</b>
<b>3</b>	<b>Submitted to University of London External System</b> Student Paper	<b>1%</b>
<b>4</b>	<b>jurnal.umk.ac.id</b> Internet Source	<b>1%</b>
<b>5</b>	<b>www.tandfonline.com</b> Internet Source	<b>1%</b>
<b>6</b>	<b>www.theseus.fi</b> Internet Source	<b>1%</b>
<b>7</b>	<b>ejournal.nusamandiri.ac.id</b> Internet Source	<b>1%</b>

[e-journal.uajy.ac.id](http://e-journal.uajy.ac.id)



# **IMPLEMENTASI ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES TERHADAP KELUHAN SARANA DAN PRASARANA DI BIRO ADMINISTRASI UMUM UNIVERSITAS MUHAMMADIYAH SURAKARTA**

## **Abstrak**

Berpendapat merupakan hak hidup manusia yang ada sejak dilahirkan, tetapi masih banyak masyarakat yang tidak menyadari kebebasan berpendapat harus diimbangi dengan tutur kata yang tidak menyakiti dan sopan agar pendapat tersampaikan secara jelas. Perkembangan teknologi saat ini memungkinkan manusia bebas berpendapat dimana saja tidak terkecuali di media sosial. *Whatsapp* merupakan salah satu media sosial berbagi pesan yang memungkinkan pengguna menyampaikan pendapat, kritik, saran dan keluhan secara personal atau dalam sebuah kelompok. Pemanfaatan pendapat untuk membantu evaluasi kinerja membutuhkan teknik analisis yang tepat sesuai dengan data yang dihadapi seperti analisis sentimen. Oleh sebab itu, penelitian ini melakukan analisis sentimen terhadap keluhan yang ada di *group whatsapp* Sarana dan Prasarana di Biro Administrasi Umum Universitas Muhammadiyah Surakarta dengan menggunakan Naïve Bayes Classifier untuk menjadi bahan masukan terhadap pengelolaan Sarpras di Universitas Muhammadiyah Surakarta. Penelitian menggunakan 420 data dengan 216 data berlabel negatif dan 204 data berlabel positif menghasilkan sentimen negatif sebesar 70,63% dan sentimen positif sebesar 29,37% dengan akurasi sebesar 69.05%. Hasil tersebut menandakan bahwa keluhan yang ada lebih banyak keluhan dengan sindiran dan ingin didahulukan daripada apresiasi kinerja.

**Kata Kunci :** Analisis Sentimen, Naïve Bayes Classifier,Whatsapp

## **Abstrak**

The opinion is the right of human life that existed since birth, but there are still many people who do not realize the freedom of opinion should be balanced with a speech that does not hurt and polite to convey the opinion clearly. Technology development allows people to freely argue that it is no exception on social media. Whatsapp is one of the social media sharing messages that allows users to personally or in a group to convey the opinion, criticism, and suggestion. To utilization opinions to help performance evaluation requires precise analytical techniques in accordance with the data encountered such as sentiment analysis. Therefore, this research conducted a sentiment analysis of complaints in WhatsApp group facilities and infrastructure General Administration Bureau University of Muhammadiyah Surakarta using Naïve Bayes Classifier to suggest BAU increase their performance manage Sarpras in University of Muhammadiyah Surakarta. This research using 420 data with 216 negative labeled data and 204 positive labeled data and resulted in negative sentiment is 70,64% and positive sentiment is 29,37% with an accuracy 69.05%. These result reveal complaints of more satire then performance appreciation.

**Keywords :** Sentiment Analysis, Naïve Bayes Classifier, Whatsapp

## 1. PENDAHULUAN

Kebebasan menyampaikan pendapat, kritik, saran dan keluhan merupakan salah satu hak asasi manusia dan merupakan kebutuhan bagi setiap masyarakat dengan syarat masih menjaga tata bahasa mengikuti norma yang ada pada masyarakat, tetapi banyak masyarakat Indonesia yang masih belum menyadari bahwa kritik, saran, serta keluhan masih banyak diutarakan dengan kata - kata sindiran, tutur kata tidak baik. Hal ini menyebabkan aspirasi yang diutarakan tidak tersampaikan dengan jelas dan hanya tutur kata tidak baik yang terungkap. Perkembangan teknologi yang sangat cepat menjadikan manusia bebas berpendapat dimanapun dan kapanpun tidak terkecuali di media sosial. *Whatsapp* merupakan salah satu media sosial berbagi pesan yang memungkinkan pengguna bertukar pesan, pendapat, kritik, dan keluhan secara individu atau dalam sebuah *group*.

*Text mining* adalah sebuah proses yang digunakan untuk mengungkapkan atau mendeteksi sebuah informasi atau pengetahuan yang terdapat pada data tekstual tidak berstruktur (Younis, 2015). Analisis sentimen merupakan bagian dari proses *text mining* yang berguna untuk mendeteksi informasi dari sebuah teks dengan hasil akhir berupa polaritas (Mandowara & Jain, 2016). Analisis sentimen bertujuan untuk memperkirakan sebuah data merupakan opini positif atau opini negatif (Thamrin & Pamungkas, 2017). Seperti penelitian dari (Fanissa, Fauzi, & Adinugroho, 2018) menentukan polaritas negatif dan positif terhadap pemilihan gubernur provinsi Sulawesi Selatan tahun 2018 dengan menggunakan dataset yang diperoleh dari *twitter*.

Analisis sentimen tidak terlepas dari sebuah algoritma *machine learning* yang digunakan untuk membentuk model berdasarkan teks untuk mengekstraksi informasi yang ada dan menentukan polaritas yang terkandung pada teks tersebut (Younis, 2015). Metode Naïve Bayes Classifier merupakan salah satu dari sekian banyak algoritma yang dipakai untuk analisis sentimen, hal ini disebabkan karena Naïve Bayes Classifier merupakan salah satu algoritma yang menggunakan metode klasifikasi dengan tingkat akurasi dan performa yang mengesankan untuk beberapa kondisi pengaplikasian di dunia nyata (Zhang & Gao, 2011). Penelitian yang dilakukan oleh (Afshoh, 2017) yang meneliti reaksi masyarakat terhadap kenaikan harga jual rokok di media sosial *twitter*



menggunakan Naïve Bayes Classifier, menunjukan bahwa algoritma Naïve Bayes efektif untuk *text mining* dikarenakan ketika dikomparasi dengan algoritma Lexicon Based tingkat akurasi Naïve Bayes lebih baik daripada Lexicon Based. Oleh karena itu, penelitian ini melakukan analisa sentimen yang ada di *group whatsapp* Sarana dan Prasarana (Sarpras) Biro Administrasi Umum (BAU) menggunakan Naïve Bayes. Penelitian ini bertujuan untuk mengetahui keluhan yang ada pada *group whatsapp* tersebut lebih banyak keluhan yang bersifat negatif atau positif (apresiasi), sehingga bisa menjadi bahan masukan terhadap kinerja BAU untuk mengelola semua hal yang berkaitan dengan sarana dan prasarana yang ada di lingkungan Universitas Muhammadiyah Surakarta. Analisa dilakukan dengan bahasa pemrograman *Python* dan *Django* untuk sistem berbasis *website* untuk mempermudah BAU melihat hasil prediksi yang sudah dilakukan.

## 2. METODE

Berdasarkan penelitian terdahulu serta data yang pada *group whatsapp* Sarpras BAU Universitas Muhammadiyah Surakarta, maka penelitian ini menggunakan Naïve Bayes Classifier untuk pemodelan data untuk mendapatkan hasil prediksi dari data *group whatsapp* Universitas Muhammadiyah Surakarta. Naïve Bayes Classifier merupakan suatu konsep algoritma untuk penentuan kelas terhadap suatu dokumen dan merupakan salah satu algoritma klasifikasi (*supervised learning*) melalui pelabelan yang mampu mengolah data dalam jumlah besar dengan tingkat akurasi yang tinggi. Persamaan umum pada algoritma Naïve Bayes dapat ditulis pada Persamaan 1 (Lestari, Perdana, & Fauzi, 2017):

$$P(L_j|N_i) = \frac{P(L_j) \times P(N_i|L_j)}{P(N_i)} \quad (1)$$

Berdasarkan persamaan diatas didapatkan bahwa  $P(L_j | N_i)$  merupakan kemunculan yang ada pada kategori j ketika terdapat kemunculan kata i (*posterior*),  $P(L_j)$  merupakan sebuah peluang ketika kemunculan dokumen yang terdapat pada kategori j,  $P(N_i|L_j)$  merupakan peluang sebuah kata i yang merupakan bagian dari kategori j (*conditional probability/likelihood*),  $P(N_i)$  merupakan peluang munculnya

sebuah kata (*evidence*),  $i$  merupakan indeks kata yang berawal dari 1 sampai kata ke- $k$ ,  $j$  merupakan indeks dari kategori yang berawal dari 1 sampai kata ke- $n$ . Penyederhanaan dapat dilakukan pada persamaan 1 dikarenakan pada perhitungan *posterior* yang ada pada proses klasifikasi peluang kemunculan kata dihilangkan dikarenakan tidak berpengaruh terhadap hasil klasifikasi tiap kategori, sehingga dituliskan dalam Persamaan 2 (Manning, Raghavan, & Schutze, 2008) :

$$P(L_j | N_i) = P(L_j) \times (N_i | L_j) \quad (2)$$

Persamaan yang digunakan dalam menghitung sebuah kemunculan sebuah dokumen dalam sebuah *Prior* atau sebuah kategori tertentu, dituliskan dalam persamaan 3 :

$$P(L_j) = \frac{NL_j}{N} \quad (3)$$

Berdasarkan persamaan diatas didapatkan bahwa  $NL_j$  merupakan jumlah dokumen latih yang masuk ke dalam kategori  $L_j$  dan  $N$  merupakan jumlah dokumen keseluruhan. Pada data yang ada pada *group whatsapp* Sarpras di Bau Universitas Muhammadiyah Surakarta perhitungan peluang kemunculan kata  $i$  yang masuk ke dalam kategori  $j$  dilakukan dengan menggunakan Multinomial Naïve Bayes, dituliskan dalam persamaan 4 :

$$P(L_j | N_i) = P(L_j) \times P(N_i | L_j) \times \dots \times P(N_k | L_j) \quad (4)$$

Multinomial Naïve Bayes merupakan salah satu dari beberapa tipe Naïve Bayes Classifier yang digunakan untuk melengkapi klasifikasi dokumen atau teks. Multinomial Naïve Bayes juga menggunakan distribusi multinomial dengan jumlah kata yang muncul sebagai fitur klasifikasi. Multinomial Naïve Bayes tidak melakukan perhitungan urutan kata dan informasi yang terdapat dalam dokumen melainkan dengan melakukan perhitungan peluang pada kata  $i$  yang terdapat pada kategori  $j$  dituliskan oleh persamaan 5 (Artissa, Asror, & Faraby, 2019). Persamaan 5 merupakan hasil dari total perhitungan *likelihood/conditional probability* data uji ditambahkan dengan angka satu untuk menghindari nilai nol kemudian dibagi dengan total perhitungan *likelihood/conditional probability* seluruh data yang ada pada kategori tertentu dan ditambah dengan kata unik yang ada pada seluruh kategori.

$$P(L_j|N_i) = \frac{\text{count}(N_i|L_j) + 1}{\text{count}(N, L_j) + |V|} \quad (5)$$

Berdasarkan persamaan diatas didapatkan bahwa  $P(L_j|N_i)$  merupakan *likelihood*,  $\text{count}(N_i|L_j) + 1$  merupakan total dari kemunculan seluruh kata yang ada pada kategori  $L_j$ ,  $|V|$  merupakan total kata unik atau kosa kata yang ada pada sebuah kategori. Penelitian analisis sentimen terhadap keluhan Sarpras BAU Universitas Muhammadiyah Surakarta menggunakan tiga fase yaitu :

1. Fase pengumpulan dataset

Diagram alir pada gambar 1 merupakan proses dari pengumpulan dataset, adapun penjelasan diagram alir tersebut sebagai berikut :

- 1) Pengumpulan dataset

Dataset dikumpulkan dari hasil konversi *chat group whatsapp* Sarpras BAU Universitas Muhammadiyah Surakarta menjadi sebuah file berformat txt lalu dikonversikan secara manual dari jangka waktu 28 Agustus 2019 sampai 22 Desember 2019 menjadi sebuah file berformat csv. Hasil konversi tersebut menghasilkan data berupa teks sebesar 420 data.

- 2) Pelabelan

Proses *labelling* atau pelabelan digunakan pada penggunaan algoritma *supervised learning* untuk mengetahui entitas tersebut memiliki label yang berbeda. Proses pelabelan manual dilakukan berdasarkan penelitian dari (Saif, Fernandez, He, & Alani, 2013) yang melakukan pelabelan manual terhadap 3000 dataset dengan menanyakan kepada 3 siswa yang sudah lulus. Proses pelabelan menggunakan 0 sebagai data yang berlabel positif dan 1 sebagai data yang berlabel negatif. Dataset *group whatsapp* Sarpras BAU Universitas Muhammadiyah Surakarta memiliki 216 kalimat yang berlabel negatif dan 204 kalimat yang berlabel positif dari total 420 data. Kalimat bisa dikatakan negatif jika mengandung unsur sindiran, ingin didahulukan, tidak sopan, sedangkan

kalimat bisa dikatakan positif jika mengandung unsur apresiasi terhadap kinerja, keluhan dengan kata yang sopan, serta motivasi terhadap peningkatan kinerja bidang Sarpras BAU Universitas Muhammadiyah Surakarta. Proses pelabelan menggunakan dua sudut pandang yaitu sudut pandang karyawan Sarpras BAU Universitas Muhammadiyah Surakarta serta sudut pandang penulis, hal ini dilakukan untuk memperoleh hasil pelabelan yang maksimal.

### 3) *Data preprocessing*

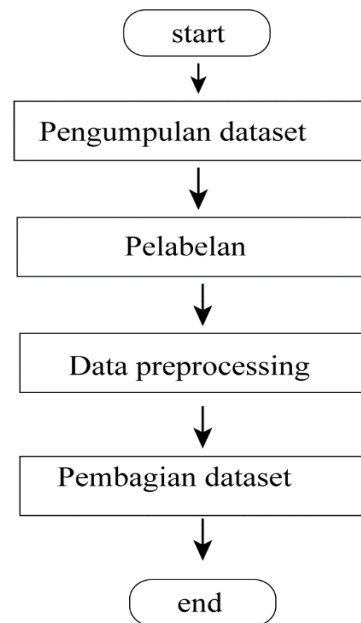
Tahap data *preprocessing* merupakan tahapan awal yang digunakan untuk mengolah kalimat yang ada menjadi sebuah analisa sentimen.

Tahapan yang digunakan dalam tahap *preprocessing* seperti :

- a. *Lowercase* merupakan proses untuk menyeragamkan sebuah teks menjadi huruf kecil seperti kalimat “Mohon Maaf, AC rusak” menjadi “mohon maaf,ac rusak”.
- b. *Tokenizing* merupakan proses yang dilakukan untuk membagi sebuah kalimat kedalam sebuah token menggunakan *library nltk* seperti kalimat “Mohon maaf, ini air di masjid sudalmiyah mati, Terima Kasih” menjadi “[“Mohon”,”maaf”,”,”,”ini”,”air”,”di”,”masjid”,”sudalmiyah”,”mati”,”,”,”Terima”,”Kasih”]”.
- c. *Data Cleansing* merupakan sebuah proses untuk pembersihan teks dari tanda baca dan karakter khusus yang ada di dalam dataset seperti kalimat “Tolong dong gedung H air,ac, lantai rusak semua, Terima Kasih!!” menjadi “Tolong dong gedung H air ac lantai rusak semua Terima Kasih”.
- d. *Data Filtering* merupakan proses untuk menghilangkan kata yang memiliki kemunculan besar, tetapi tidak bermakna (*stopwords*) seperti kalimat “Gedung FKI lantai kamar mandi dan kran air rusak” menjadi “Gedung FKI lantai kamar mandi kran air rusak” serta mengembalikan kata menjadi kata dasar



dengan menghilangkan semua imbuhan kata meliputi awalan kata, sisipan kata dan akhiran kata seperti kalimat “Mahasiswa farmasi terjatuh ketika melewati lantai licin” menjadi “Mahasiswa farmasi jatuh ketika lewat lantai licin”. Data *Filtering* menggunakan *library nltk* untuk menghilangkan kata tidak bermakna dan *python sastrawi* untuk mengembalikan kata menjadi kata dasar.



Gambar 1. Diagram alir pengumpulan dataset

## 2. Fase pembuatan sistem

Fase pembuatan sistem dilakukan untuk memudahkan pihak Sarpras BAU Universitas Muhammadiyah Surakarta mengetahui hasil prediksi dengan jelas. Pembuatan sistem menggunakan spesifikasi perangkat keras, perangkat lunak serta *library* sebagai berikut :

- a) Perangkat keras menggunakan laptop dengan spesifikasi *Processor* Intel i5 6200U, RAM 8 GB, Harddisk 1 TB, Linux Ubuntu 18.04 LTS.
- b) Perangkat lunak yang digunakan diantaranya adalah Visual Studio Code, Google Chrome, Jupyter Notebook, dan Anaconda.

- c) *Library* yang digunakan diantaranya adalah *pandas*, *nltk*, *sklearn*, *highchart*.

Pengembangan sistem menggunakan bahasa pemrograman *Python* dan *Django* untuk mengkonversikan menjadi sebuah sistem berbasis *website*, *Bootstrap* untuk membuat tampilan antarmuka serta menggunakan *MySQL* untuk menyimpan data. *MySQL* merupakan salah satu *software* yang digunakan sebagai *database* tempat penyimpanan sebuah data, struktur *database* yang digunakan dalam sistem ini terdiri dari dua tabel yaitu, tabel dataset yang terdiri dari dua atribut yakni atribut tahun sebagai *primary key* dan atribut data *training*, sedangkan untuk tabel hasil prediksi terdiri dari 4 atribut yakni atribut tahun sebagai *primary key*, atribut sentimen negatif, atribut sentimen positif dan atribut *accuracy*, detail tabel bisa dilihat pada gambar 2.



Dataset
year_2: VARCHAR(200)
data_training: VARCHAR(200)

Hasil_prediksi
year_2: VARCHAR(200)
sentiment_negatif: VARCHAR(200)
sentiment_positif: VARCHAR(200)
accuracy: VARCHAR(100)

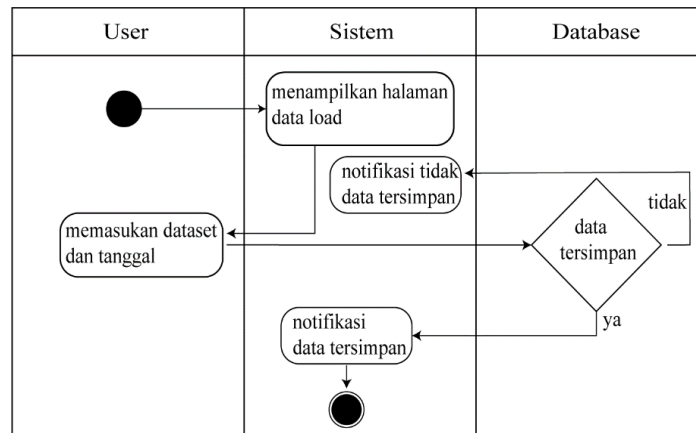
Gambar 2. Tabel *database*

Sistem mempunyai tiga halaman pemrosesan mulai dari memasukan data, menganalisis data dan visualisasi data, adapun tiga halaman tersebut yaitu:

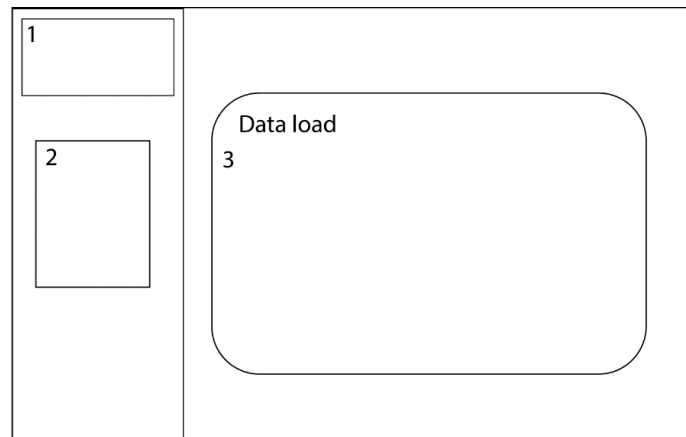
1) Halaman data *load*

Gambar 3 merupakan diagram aktivitas yang ada di dalam halaman data *load*. Pengguna akan memasukan dataset berformat csv serta tanggal dataset diperoleh dengan *form* yang sudah tersedia, jika data tersebut berhasil dimasukan maka akan ada notifikasi pemberitahuan data tersebut sudah tersimpan sedangkan jika data tersebut tidak tersimpan di dalam *database* maka sistem akan memberikan notifikasi pemberitahuan data tersebut tidak tersimpan. Gambar 4 merupakan perancangan halaman data *load*, adapun nomor 1 menunjukkan tempat logo *website*, nomor 2 menunjukkan *navigation bar* yang ada di dalam

*website*, dan nomor 3 merupakan *form* untuk pengguna memasukan dataset serta tanggal.



Gambar 3. Diagram aktivitas halaman data load

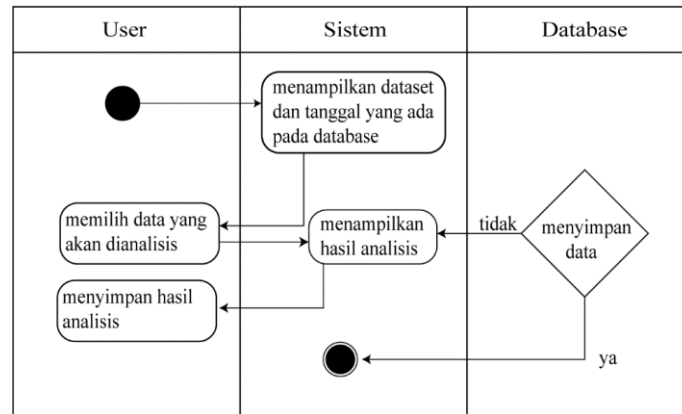


Gambar 4. *Mockup* halaman data load

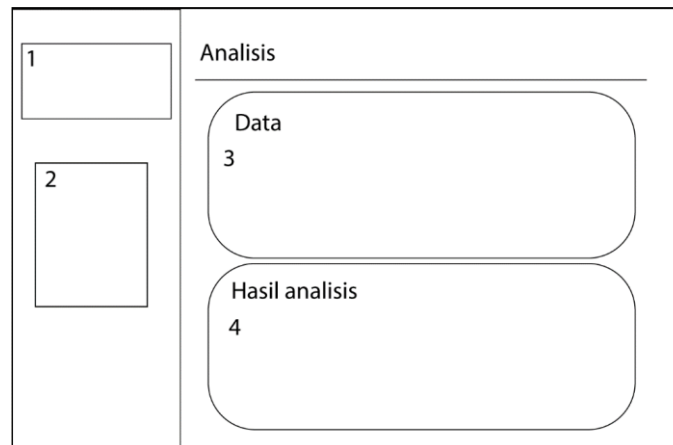
## 2) Halaman analisis

Gambar 5 merupakan diagram aktivitas yang ada di dalam halaman analisis. Pengguna memilih data yang tersedia di halaman analisis, setelah memilih data sistem akan menganalisis data dengan membaca dataset tersebut menggunakan *library pandas* tersebut dan mengkonversikannya menjadi presentasi negatif, presentasi positif serta akurasi. Hasil analisis merupakan bagian dari pemodelan data yang menggunakan *library sklearn* dengan algoritma Naïve Bayes Classifier tipe Multinomial Naïve Bayes, ketika hasil analisis sudah ditampilkan maka pengguna menyimpan hasil analisis tersebut di dalam *database*.

Gambar 6 merupakan perancangan halaman analisis, adapun nomor 1 merupakan logo *website*, nomor 2 merupakan *navigation bar* yang ada *website*, nomor 3 merupakan data yang sudah tersimpan di dalam *database*, nomor 4 merupakan hasil analisis dari data yang sudah dipilih pengguna.



Gambar 7. Diagram aktivitas halaman analisis



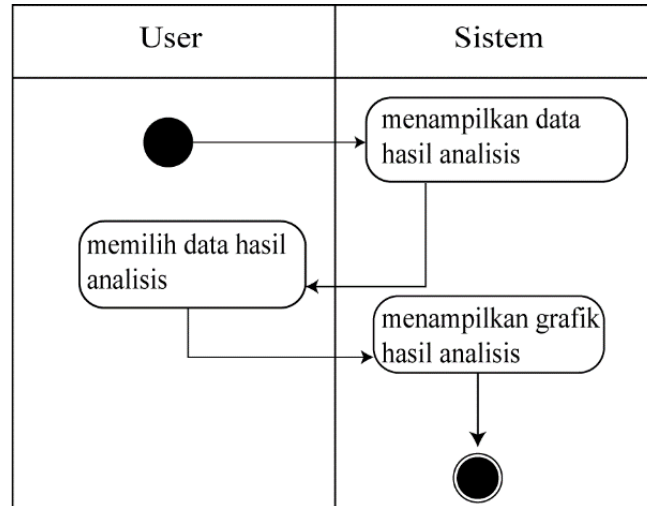
Gambar 6. *Mockup* halaman analisis

### 3) Halaman grafik

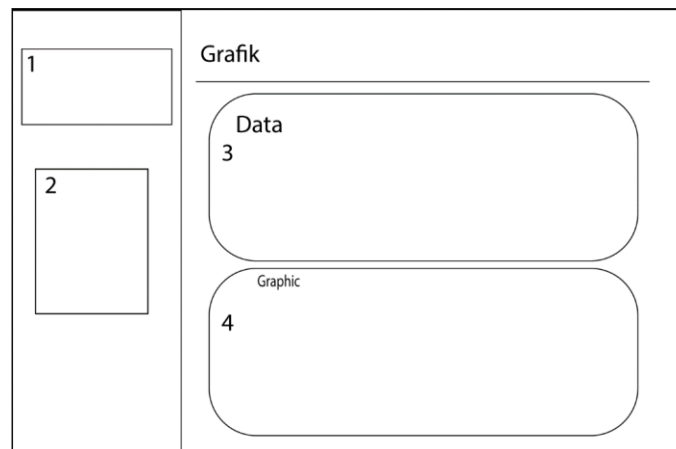
Gambar 7 merupakan diagram aktivitas yang ada di dalam halaman grafik. Pengguna memilih hasil analisis yang ada pada halaman grafik, setelah pengguna memilih hasil analisis maka sistem akan mengkonversikan hasil analisis tersebut ke dalam grafik. Perancangan antarmuka halaman grafik ditunjukkan pada gambar 8, adapun nomor 1 merupakan logo *website*, nomor 2 merupakan *navigation bar*, nomor 3



merupakan data hasil analisis yang ada di dalam *database*, dan nomor 4 merupakan grafik dari data hasil analisis yang sudah dipilih oleh pengguna.



Gambar 7. Diagram aktivitas halaman grafik



Gambar 8. *Mockup* halaman grafik

### 3. Pengujian performa

Pengujian performa digunakan untuk mengukur keberhasilan sebuah algoritma dalam sebuah pemodelan, pengujian performa menggunakan Jupyter Notebook dan menggunakan *library sklearn* untuk pengujian serta *library matplotlib* untuk visualisasi hasil pengujian. *Confusion matrix* merupakan sebuah pengukuran kinerja algoritma, matrix ini menunjukkan hubungan antara benar dan salah dalam memprediksi sebuah label. Atribut

*Confusion matrix* terdiri dari TP (*True Positive*) mewakili jumlah label positif yang benar dalam klasifikasi sentimen, FP (*False Positive*) mewakili jumlah label positif tetapi diprediksi masuk ke dalam label negatif, TN (*True Negative*) mewakili jumlah label negatif yang benar dalam klasifikasi sentimen, FN (*False Negative*) mewakili jumlah label negatif yang diprediksi masuk ke dalam label positif (Tripathy, Agrawal, & Rath, 2015). Detail atribut *confusion matrix* dapat dilihat pada tabel 1.

Tabel 1. *Confusion matrix*

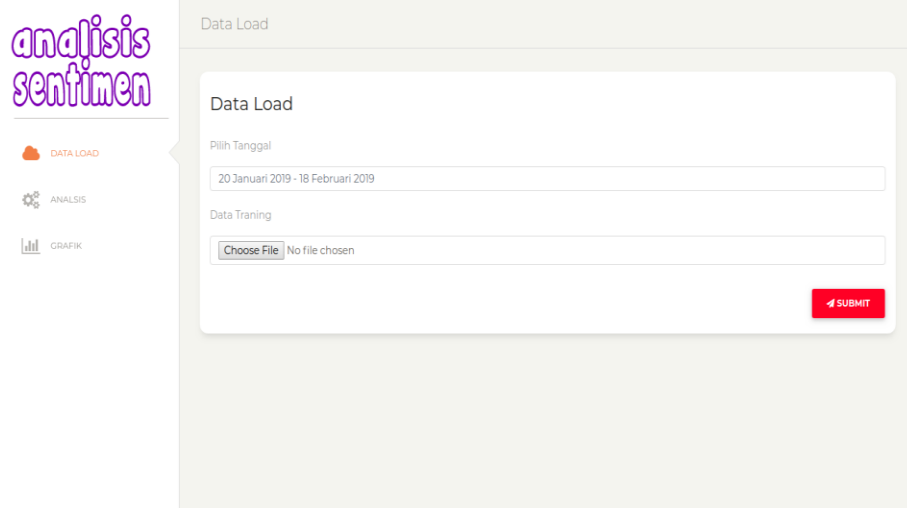
	Label	
	Positif	Negatif
Positif	<i>True Positive</i>	<i>False Positive</i>
Negatif	<i>False Negative</i>	<i>True Negative</i>

Berdasarkan tabel tersebut didapatkan beberapa indikator performa seperti *precision*, *recall*, *f1-score* serta *accuracy* dapat dilakukan berdasarkan tabel *confusion matrix*. *Precision* merupakan rasio jumlah yang benar diprediksi positif untuk jumlah label positif, *recall* merupakan rasio jumlah yang diprediksi benar terhadap keseluruhan data yang benar, *f1-score* merupakan rata – rata dari *precision* dan *recall*, sedangkan akurasi merupakan nilai keseluruhan prediksi yang benar.

### 3. HASIL DAN PEMBAHASAN

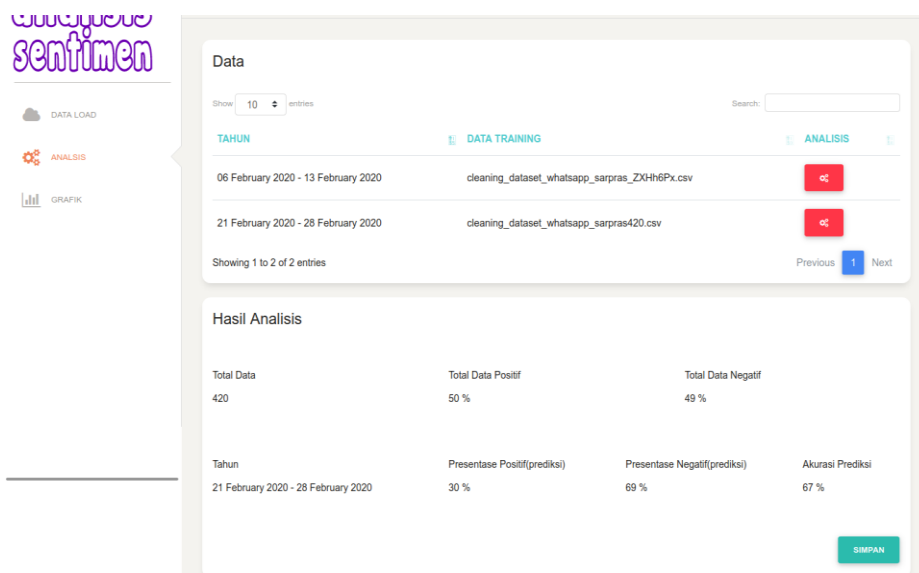
#### 3.1 Hasil

Gambar 9 merupakan hasil halaman data *load* yang digunakan untuk *input* dataset beserta tanggal kedalam *database*.



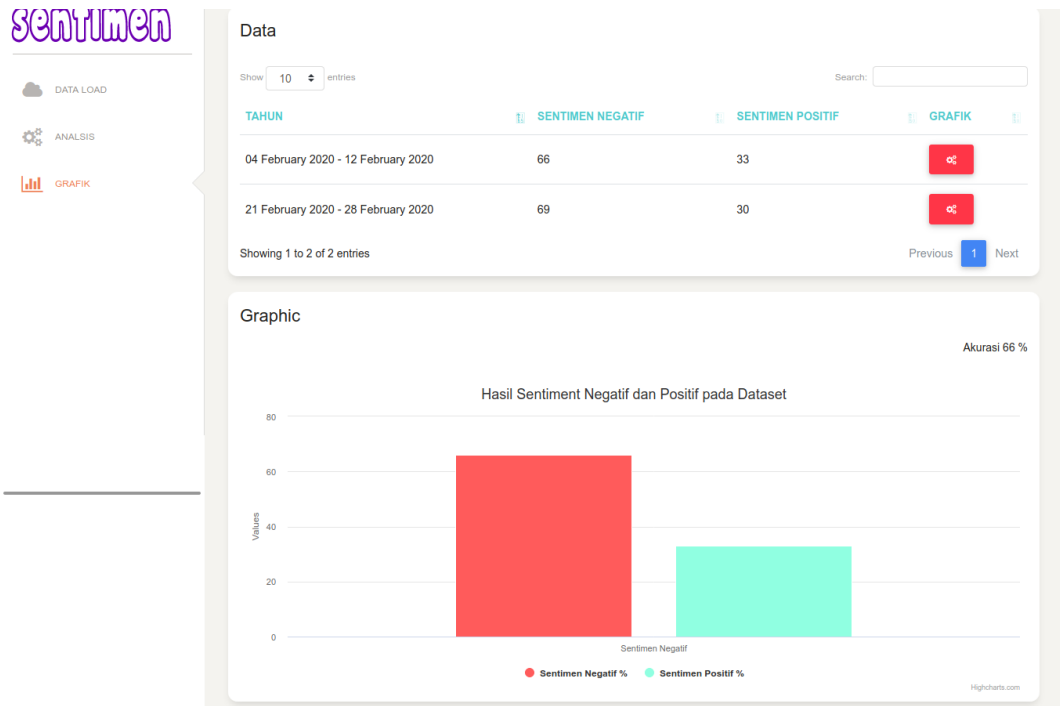
Gambar 9. Halaman data *load*

Gambar 10 merupakan halaman analisis yang digunakan untuk menganalisis dataset yang sudah ada sesuai dengan pilihan pengguna.



Gambar 10. Halaman hasil analisis

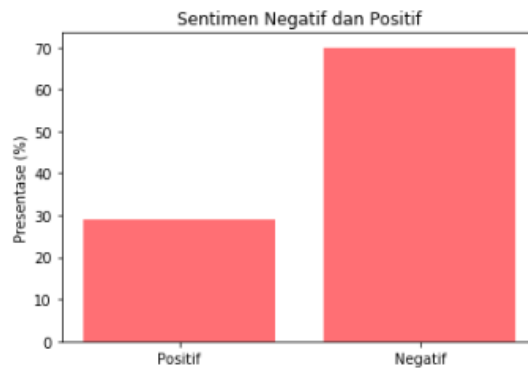
Gambar 11 merupakan hasil halaman grafik digunakan untuk memberikan visualisasi presentase negatif dan positif dari hasil pemodelan dataset.



Gambar 11. Halaman Grafik

### 3.2 Hasil Analisis

Hasil analisis menggunakan Naïve Bayes Classifier dengan dataset sebesar 420 buah dengan data latih sebesar 70% dan data uji sebesar 30% diambil secara acak dari total keseluruhan dataset menghasilkan presentase label positif sebesar 29,37% sedangkan label negatif sebesar 70,63% seperti pada gambar 12. Akurasi prediksi sebesar 69.05% dengan menggunakan *library sklearn*.



Gambar 12. Gambar presentase hasil prediksi dataset



Akurasi tersebut menandakan bahwa pemodelan data dari proses pelabelan sampai pemodelan cukup berhasil. Nilai *precision* lebih besar pada data yang berlabel positif sebesar 48 % dibanding nilai *precision* pada data yang berlabel negatif sebesar 92 %. Ketimpangan hasil *precision* yang dapat dilihat pada tabel 2 dikarenakan data latih lebih banyak data yang berlabel positif serta pemilihan algoritma sangat mempengaruhi dikarenakan Multinomial Naïve Bayes menghitung jumlah kata yang ada pada label.

Tabel 2. Hasil Pengujian

Label	Naïve Bayes		
	Precision	Recall	F1 Score
Positif	48 %	86 %	62 %
Negatif	92 %	62 %	74 %

Berdasarkan hasil penelitian diatas didapatkan bahwa *group whatsapp* Sarpras BAU Universitas Muhammadiyah Surakarta lebih banyak opini negatif dibandingkan dengan opini positif.

### 3.3 Pembahasan

Penelitian tentang mengetahui sentimen yang ada pada *group whatsapp* Sarpras BAU Universitas Muhammadiyah Surakarta dari tahap pelabelan sampai pemodelan memiliki beberapa permasalahan yaitu :

#### 3.3.1 Kendala Memahami Opini

Kendala memahami opini terjadi dikarenakan di dalam sebuah kalimat mengandung beberapa makna seperti keluhan, sindiran, serta apresiasi.

#### 3.3.2 Kendala pada Proses Pelabelan

Kendala yang sering kali ditemui pada saat pelabelan adalah terdapat beberapa dataset yang di awal kalimat mengandung kata berkategori positif tetapi di tengah kalimat mengandung kata berkategori negatif. Hal ini menyebabkan sebuah kalimat bersifat ambigu antara opini apresiasi atau opini keluhan yang menyebabkan kesalahan pada proses pelabelan, seperti kalimat

“Mohon maaf, AC di gedung H mati lantai rusak semua orang kepanasan sudah dionlinekan tetapi tidak ada respon”. Kalimat tersebut merupakan kalimat yang berkategori aduan dan sindiran berkategori negatif dan mengandung kata yang ada pada kalimat berkategori aduan halus seperti kata “Mohon maaf”. Ketika dilakukan pengujian pada sistem kalimat yang bersifat aduan halus seperti “Mohon maaf, Pak air di gedung H sedang mati sudah dionlinekan, Terima Kasih”. Kalimat tersebut merupakan kalimat yang bersifat aduan halus tanpa ada sindiran atau ingin didahulukan, tetapi ketika dilakukan pengujian di sistem menghasilkan kalimat yang berlabel negatif diakibatkan ada dua suku kata yang ada pada kalimat negatif. Hal ini dibuktikan dengan kata yang sering muncul di label positif yaitu, terima kasih sebesar 13 kemunculan, tim sebesar 10 kemunculan, gedung sebesar 10 kemunculan, dan mohon sebesar 14 kemunculan sedangkan kata yang sering muncul di label negatif yaitu, gedung sebesar 23 kemunculan, mohon sebesar 14 kemunculan, rusak sebesar 9 kemunculan, dan mati sebesar 12 kemunculan.

### **3.3.3 Ketidakseimbangan Data**

Ketidakseimbangan data merupakan penyebab dari kendala memahami opini dan kendala dalam proses pelabelan, walaupun selisih data yang berlabel positif dan berlabel negatif tidak berbeda jauh hanya selisih 12 data tetapi sangat berdampak pada hasil analisis.

## **4. PENUTUP**

Berdasarkan penelitian yang sudah dilakukan dapat disimpulkan bahwa :

1. Ketidakseimbangan data memberikan pengaruh terhadap *precision* yang mengalami ketimpangan, hal ini dikarenakan Naïve Bayes Classifier dengan tipe Multinomial Naïve Bayes memperhitungkan jumlah *term* atau kata yang ada pada sebuah label.
2. Penelitian dengan 420 data yang terdiri dari 216 data berlabel negatif dan 204 data berlabel positif yang ada di *group whatsapp* Sarpras BAU

Universitas Muhammadiyah Surakarta lebih banyak mengandung sentimen negatif yang berisi opini keluhan dengan jenis ingin didahulukan dan sindiran sebesar 70,63% sedangkan opini apresiasi sebesar 29,37% setelah pemodelan dataset.

3. Sistem mempunyai kelemahan dikarenakan pengumpulan dataset dan pelabelan masih manual.

#### **DAFTAR PUSTAKA**

- Afshoh, F. (2017). *Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter*. Universitas Muhammadiyah Surakarta.
- Artissa, Y. B. N. D., Asror, I., & Faraby, S. A. (2019). Personality Classification based on Facebook status text using Multinomial Naïve Bayes method. *Journal of Physics: Conference Series*, 1192(1).
- Fanissa, S., Fauzi, M. A., & Adinugroho, S. (2018). *Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking*. 2(8), 2766–2770.
- Lestari, A. R. T., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1718–1724.
- Mandowara, J., & Jain, A. (2016). Text Classification by Combining Text Classifiers to Improve the Efficiency of Classification. *International Journal of Computer Application*, 6(2), 126–129.
- Manning, C. D., Raghavan, P., & Schutze, H. (2008). Text classification and Naive Bayes. In *Introduction to Information Retrieval* (pp. 234–265).
- Saif, H., Fernandez, M., He, Y., & Alani, H. (2013). Evaluation datasets for Twitter sentiment analysis : a survey and a new dataset , the STS-Gold. *1st Interantional Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013)*.
- Thamrin, H., & Pamungkas, E. W. (2017). A Rule Based SWOT Analysis Application :

- A Case Study for Indonesian Higher Education Institution. *Procedia Computer Science*, 116, 144–150.
- Tripathy, A., Agrawal, A., & Rath, S. K. (2015). Classification of Sentimental Reviews Using Machine Learning Techniques. *Procedia Computer Science*, 57, 821–829.
- Younis, E. M. G. (2015). Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study. *International Journal of Computer Applications*, 112(5), 44–48.
- Zhang, W., & Gao, F. (2011). An Improvement to Naive Bayes for Text Classification. *Procedia Engineering*, 15, 2160–2164.